The authors did a great job to improve the paper. However, there are still some needed modifications. This should be done for the following questions. I recommend that the authors answer the questions not only in the body of the manuscript but also in the reply to reviewer files, not just mentioning “e.g., Line 316”. Added to this, I just looked at my questions:

We thank the reviewers for their analysis of the manuscript and the previous reviews.

1. The clustering approach is not clear. For example, clustering data into three groups, how does this approach help in modeling?

The reviewer refers to CA-1. We adapted the corresponding text section in lines 151-154 to add further information with this regards: “in CA-1, k-means clustering was used to split the soil texture data of both depth levels into three clusters to distinguish between sandy, silty and clayey soils (Fig. 2). The distribution of every predictor value among the three clusters is analysed, to a) determine whether the predictor has any influence on soil texture (feature selection in chapter 2.2.3), and b) gain process understanding by analysing the relationship between predictors and soil texture.”

2. Feature selection is also unclear. Did you specify the significant relationships between soil fractions and covariates? If yes, this is not a Filter approach?!

According to our understanding, filter methods are a broad term for procedures which quantify the individual predictor-response relationship, and define a threshold to “filter”-out weak predictors. One of the more popular filter methods is ANOVA. Kruskal–Wallis is a nonparametric equivalent to ANOVA.

3. Did you transform soil fractions? If you use the raw data, the predicted maps do not guarantee the sum of 100%. You should transform data and then do modeling.

The respective contributions of the particle size fractions of the legacy soil data sum up to 100%. Still, training models from these data may result in predictions which do not sum up to 100%. We have addressed this issue in lines 311-315: "It needs to be noted that the maps of predicted values are showing the results of independent models for different soil texture classes, and the results don’t add up to 100%. The method to scale the data to 100% should be selected with the purpose of the specific data utilization in mind. Different approaches could include leaving one of the texture classes out and summing up to 100%, weighted scaling by texture class or weighted scaling by the regional accuracy of the texture classes."

4. Random oversampling is just “copy and paste” of the original data. Do not think the approach resulted in overfitting? I mean that Random oversampling is not a good approach to balance the datasets. Please use another approach.

Oversampling can be used as a tool to shift the model along the variance-bias gradient: If it is assumed that the machine learning model has a better representation of the majority class, the overall model error can be increased while reducing the error in the minority samples. Please compare Taghizadeh-Mehrjardi et al. (2019), and Sharififar et al. (2019).

5. Please calculate line concordance instead of R squared. I mean R squared is not a good indicator to show the accuracy of the models.

We decided for the use of two error metrics: a scaled RMSE and R squared. RMSE allows an estimation of the deviation between observed and modelled results, however R squared is a more

popular metric in the field, that allowed us to compare the model performance with other publications and put our results into perspective.